

UNSUPERVISED DIFFUSION-BASED LMS FOR NODE-SPECIFIC PARAMETER ESTIMATION OVER WIRELESS SENSOR NETWORKS

Jorge Plata-Chaves, Mohamad Hasan Bahari, Marc Moonen, Alexander Bertrand

Department of Electrical Engineering-ESAT, STADIUS, KU Leuven, B-3001 Leuven, Belgium
E-mails: {jplata, mohamadhasan.bahari, marc.moonen, alexander.bertrand}@esat.kuleuven.be

ABSTRACT

We study a distributed node-specific parameter estimation problem where each node in a wireless sensor network is interested in the simultaneous estimation of different vectors of parameters that can be of local interest, of common interest to a subset of nodes, or of global interest to the whole network. We assume a setting where the nodes do not know which other nodes share the same estimation interests. First, we conduct a theoretical analysis on the asymptotic bias that results in case the nodes blindly process all the local estimates of all their neighbors to solve their own node-specific parameter estimation problem. Next, we propose an unsupervised diffusion-based LMS algorithm that allows each node to obtain unbiased estimates of its node-specific vector of parameters by continuously identifying which of the neighboring local estimates correspond to each of its own estimation tasks. Finally, simulation experiments illustrate the efficiency of the proposed strategy.

Index Terms— Distributed node-specific parameter estimation, wireless sensor networks, diffusion algorithm, adaptive clustering.

1. INTRODUCTION

In most distributed estimation problems, it is generally assumed that the nodes in a wireless sensor network (WSN) are interested in the same network-wide signal or parameter (e.g. [1]–[5]). However, some applications such as speech enhancement in acoustic sensor networks [6]–[8], beamforming [9], or cooperative spectrum sensing in cognitive radio networks [10]–[11] are multi-task oriented. In these cases, special attention is required to more general distributed estimation techniques where the nodes cooperate although they have different but still partially-overlapping estimation interests and their observations may arise from different models.

In the growing literature on node-specific parameter estimation (NSPE) problems over adaptive WSNs, two major groups of works can be identified. The first group assumes that all nodes know a priori the relationship between their estimation tasks and the estimation tasks of their neighbors. Within this category, the aforemen-

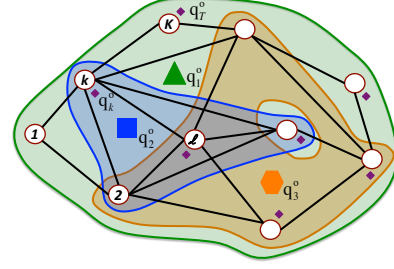


Fig. 1. A wireless network of K nodes with NSPE interests.

tioned prior information is leveraged to derive strategies that provide asymptotically unbiased solutions in an NSPE setting where the nodes have both overlapping and arbitrarily different estimation interests [12]–[16]. Additionally, this prior information is leveraged by different diffusion-based algorithms that apply different spatial regularizers to let each node solve its estimation task by using the local estimates of neighboring nodes with numerically similar estimation interests [17]–[18]. To avoid the bias that results from the combination of local estimates associated with different tasks [19]–[20], the second group of algorithms implement an inference algorithm together with an adaptive clustering technique that allows the nodes to infer which of their neighbors have the same interest [20]–[24]. However, since the proposed strategies run over diffusion networks where each node is only interested in one vector of parameters, the cooperation is finally limited to nodes that have exactly the same objectives once the clustering technique has converged.

To the best of our knowledge, there are no unsupervised strategies that solve an NSPE problem where the nodes simultaneously estimate parameter vectors of local, common and/or global interest and where there is no prior information about the relationship between the NSPE tasks. Motivated by this fact, we propose an unsupervised diffusion-based LMS for NSPE with combination coefficients determined through a multi-task clustering technique in an adaptive fashion. In this clustering technique, each node solves a hypothesis testing problem to determine which of the local estimates of its neighbors correspond to each of its estimation tasks. Unlike the existing algorithms [20]–[24], the proposed scheme can yield asymptotically unbiased solutions and allows a beneficial cooperation among the nodes although they have different interests. Indeed, due to the exponential rate of decay for the probabilities of erroneous clustering, as the computer simulations show, the proposed scheme achieves the same steady-state mean square deviation (MSD) as the diffusion-based NSPE LMS (D-NSPE) algorithm that knows a priori the relationship between the node-specific tasks.

This work was carried out at the ESAT Laboratory of KU Leuven, in the frame of KU Leuven Research Council CoE PFV/10/002 (OPTec) and BOF/STG-14-005, Concerted Research Action GOA-MaNet, the Interuniversity Attractive Poles Programme initiated by the Belgian Science Policy Office IUAP P7/23 ‘Belgian network on stochastic modeling analysis design and optimization of communication systems’ (BESTCOM) 2012-2017, Research Project FWO nr. G.0763.12 ‘Wireless Acoustic Sensor Networks for Extended Auditory Communication’, and EU/FP7 project HANDiCAMS. The project HANDiCAMS acknowledges the financial support of the Future and Emerging Technologies (FET) programme within the Seventh Framework Programme for Research of the European Commission, under FET-Open grant number: 323944. The scientific responsibility is assumed by its authors.

2. PROBLEM FORMULATION

We consider a network consisting of K nodes randomly deployed over some region. Nodes that are able to exchange information in one hop are said to be neighbors. The neighborhood of any particular node k , including node k itself, is denoted as \mathcal{N}_k . To ensure that the network is connected (see Fig. 1), the neighborhoods are set so that there is a path between any pair of nodes in the network.

In the considered network, each node k , at discrete time i , has access to data $\{d_{k,i}, U_{k,i}\}$, which are realizations related to events that coexist in the monitored area and which follow the relation

$$\mathbf{d}_{k,i} = \mathbf{U}_{k,i} w_k^o + \mathbf{v}_{k,i} \quad (1)$$

where

- w_k^o equals the vector of dimension M_k that gathers all parameters of interest for node k ,
- $\mathbf{v}_{k,i}$ denotes measurement and/or model noise with zero mean and covariance matrix $R_{v_{k,i}}$ of dimensions $L_k \times L_k$,
- $\mathbf{d}_{k,i}$ and $\mathbf{U}_{k,i}$ are zero-mean random variables with dimensions $L_k \times 1$ and $L_k \times M_k$, respectively.

By processing the data set $\{d_{k,i}, U_{k,i}\}_{i=1}^K$, the objective of the network consists in finding the node-specific estimators $\{w_k\}_{k=1}^K$ that minimize the following cost function

$$J_{\text{glob}}(\{w_k\}_{k=1}^K) = \sum_{k=1}^K E \{ \|\mathbf{d}_{k,i} - \mathbf{U}_{k,i} w_k\|^2 \}. \quad (2)$$

Unlike in papers addressing a single-task estimation problem, e.g., [3]-[4], here the parameter estimation interests may vary from node to node in (1), i.e., $w_k^o \neq w_\ell^o$ if $k \neq \ell$. Indeed, following the novel node-specific observation model considered in [14], [16], it is assumed that each node-specific vector w_k^o may consist of a sub-vector of parameters of global interest to the whole network, sub-vectors of parameters of common interest to subsets of nodes including node k , and a sub-vector of local parameters for node k . In particular, it is considered that

$$w_k^o = \text{col}\{\mathbf{q}_t^o\}_{t \in \mathcal{T}_k} \quad (3)$$

where $\text{col}\{\cdot\}$ denotes a column operator stacking arguments on top of each other and where \mathbf{q}_t^o denotes an $(M_t \times 1)$ vector of parameters associated with the global, common or local estimation task t and where \mathcal{T}_k equals an ordered set of indices t associated with the $n_k = |\mathcal{T}_k|$ vectors \mathbf{q}_t^o that are of interest for node k . It is noted that $M_k = \sum_{t \in \mathcal{T}_k} M_t$ and that $\mathcal{T}_k \in \mathcal{T}$ where \mathcal{T} is the set of all parameter estimation tasks in the entire network. As a result, the observation model in (1) can now be rewritten as

$$\mathbf{d}_{k,i} = \sum_{t \in \mathcal{T}_k} \mathbf{U}_{k_t,i} \mathbf{q}_t^o + \mathbf{v}_{k,i} \quad (4)$$

where $\mathbf{U}_{k_t,i}$, equals a matrix of dimensions $L_k \times M_t$ that consists of the columns of $\mathbf{U}_{k,i}$ associated with \mathbf{q}_t^o . From (4), also note that the considered NSPE problem can be cast as minimizing

$$\sum_{k=1}^K E \|\mathbf{d}_{k,i} - \sum_{t \in \mathcal{T}_k} \mathbf{U}_{k_t,i} \mathbf{q}_t\|^2 \quad (5)$$

with respect to variables $\{\mathbf{q}_t\}_{t \in \mathcal{T}}$.

The algorithms derived in [14] and [16] yield unbiased estimates of the node-specific vector of parameters by seeking the minimizer

of the cost function in (5) under an incremental or diffusion mode of cooperation, respectively. However, to do so, each node should know a priori which of its neighbors share the same parameter estimation interests, i.e., are interested in estimating $\{\mathbf{q}_t^o\}_{t \in \mathcal{T}_k}$. Unfortunately, this prior information is not available in many scenarios. *In this paper, we consider the more challenging problem of deriving a diffusion-based LMS algorithm that is able to solve the NSPE problem stated in (5) when the nodes do not know a priori which of their neighbors share the same NSPE interests.*

3. PERFORMANCE OF THE D-NSPE LMS IN A SETTING WITH UNKNOWN RELATIONSHIPS BETWEEN TASKS

In this section, in order to motivate the derivation of the proposed algorithm, we briefly analyze the performance of the diffusion-based solution derived in [16] when a node does not know which of the local estimates of its neighbors correspond to each of its tasks. For the sake of an easy exposition and without loss of generality, we assume that $M_t = M$ for all $t \in \mathcal{T}$.

In brief, the D-NSPE LMS algorithm [16] is able to minimize (5) by implementing the following recursion at each node k :

$$\begin{cases} \textbf{Adaptation step.} \text{ For each task } t \in \mathcal{T}_k \text{ execute} \\ \psi_{k,t}^{(i)} = \phi_{k,t}^{(i-1)} + \mu_k \mathbf{U}_{k_t,i}^H \left[\mathbf{d}_{k,i} - \sum_{p \in \mathcal{T}_k} \mathbf{U}_{k_p,i} \phi_{k,p}^{(i-1)} \right] \\ \textbf{Combination step.} \text{ For each task } t \in \mathcal{T}_k \text{ execute} \\ \phi_{k,t}^{(i)} = \sum_{\ell \in \mathcal{N}_k} \sum_{p \in \mathcal{T}_\ell} c_{k\ell,tp}(i) \psi_{\ell,p}^{(i)}. \end{cases} \quad (6)$$

where $\phi_{k,t}^{(i)}$ is the local estimate of \mathbf{q}_t^o at node k and time instant i and $\{c_{k\ell,tp}(i)\}$ are time-varying convex combination coefficients that satisfy

$$\begin{cases} c_{k\ell,tp}(i) > 0 & \text{if } \ell \in \mathcal{N}_k \cap \mathcal{C}_t \text{ and } p = t, \\ c_{k\ell,tp}(i) = 0 & \text{otherwise.} \end{cases} \quad (7)$$

and

$$\sum_{\ell \in \mathcal{N}_k} \sum_{p \in \mathcal{T}_\ell} c_{k\ell,tp}(i) = 1 \quad (8)$$

with \mathcal{C}_t denoting the set of nodes interested in estimating \mathbf{q}_t^o . Note that there are several policies to select the combination coefficients. On the one hand, the combination rule can be static, e.g., as in the uniform, Metropolis or relative-degree rule [25]. On the other hand, the coefficients can be adapted over time (e.g. see [21]).

Independently of the selected combination policy, from (6) and (7), it can be noticed that the combination step associated with the estimation of \mathbf{q}_t^o at node k can only process local estimates of the same vector of parameters, i.e., $p = t$, transmitted by neighboring nodes, i.e., $\ell \in \mathcal{N}_k$. This particular constraint on the set of possible combination policies ensures asymptotical unbiasedness [16], i.e., $\lim_{i \rightarrow \infty} E\{\tilde{\mathbf{q}}_{k,t}^{(i)}\} = 0$ with $\tilde{\mathbf{q}}_{k,t}^{(i)} = \mathbf{q}_t^o - \text{col}\{\phi_{k,t}^{(i)}\}_{t \in \mathcal{T}_k}$ and w_k^o defined in (3). However, it requires each node to know a priori which of the local estimates exchanged by its neighbors correspond to each of its parameter estimation tasks.

When the nodes do not know a priori the relationship between the tasks, each node could implement a stand-alone LMS to solve its NSPE problem, which is equivalent to the implementation of (6)-(8) with $c_{kk,tt}(i) = 1$. Although this approach would again allow to find unbiased estimates of the node-specific vector of parameters that minimize (5), the nodes cannot take advantage of the well-known benefits provided by the cooperation. Alternatively, each node could

blindly fuse all the local estimates exchanged by all its neighbors, which yields an implementation of (6) where the convex combination coefficients satisfy (8) and

$$\begin{cases} c_{k\ell, tp}(i) = \alpha_{k\ell, tp} > 0 & \text{if } \ell \in \mathcal{N}_k, \\ c_{k\ell, tp}(i) = 0 & \text{otherwise.} \end{cases} \quad (9)$$

with $\alpha_{k\ell, tp}$ equal to a positive constant for all $k \in \{1, 2, \dots, K\}$, $\ell \in \mathcal{N}_k$, $t \in \mathcal{T}_k$ and $p \in \mathcal{T}_\ell$. However, under this approach the estimates at the nodes will be biased. In particular, assuming that

- A1) $\mathbf{v}_{k,i}$ is temporally and spatially white noise that is independent of $\mathbf{U}_{\ell, i'}$ for all ℓ and i' , with $k, \ell \in \{1, 2, \dots, K\}$.
- A2) $\mathbf{U}_{k,i}$ is temporally stationary, white and spatially independent with $R_{U_k} = E\{\mathbf{U}_{k,i} \mathbf{U}_{k,i}^H\}$;
- A3) $\mathbf{U}_{k_t, i}$ and $\mathbf{U}_{k_p, i}$ are independent for all $k \in \{1, 2, \dots, K\}$ and $t \neq p$,

the asymptotic bias of the estimates resulting from this cooperative approach is given by the following theorem (the proof is omitted due to space constraints).

Theorem 1. *For any initial conditions and under assumptions A1-A3, if the positive step-size of each node satisfies*

$$\mu_k < 2/\lambda_{\max}(\{R_{U_{k_t}}\}_{t \in \mathcal{T}_k}), \quad (10)$$

then the estimates generated by the D-NSPE LMS algorithm summarized in (6) converge in the mean when the convex combination coefficients satisfy (8) and (9). Furthermore, the estimation bias in the steady-state tends to

$$E\{\tilde{\mathbf{q}}^{(\infty)}\} = [I - \check{C}[I - MD]]^{-1}[I - \check{C}]\mathbf{q}^o \quad (11)$$

with $\tilde{\mathbf{q}}^{(i)} = \text{col}\{\text{col}\{\tilde{\mathbf{q}}_{k,t}^{(i)}\}_{t \in \mathcal{T}_k}\}_{k=1}^K$, $\mathbf{q}^o = \text{col}\{\text{col}\{\mathbf{q}_t^o\}_{t \in \mathcal{T}_k}\}_{k=1}^K$, $D = \text{diag}\{R_{U_k}\}_{k=1}^K$, $M = \text{diag}\{\mu_k I_{M_k}\}_{k=1}^K$ and $\check{C} = C \otimes I_M$ where

$$C = [c_{1, \mathcal{T}_1(1)} \cdots c_{1, \mathcal{T}_1(n_1)} \quad c_{2, \mathcal{T}_2(1)} \cdots c_{K, \mathcal{T}_K(n_K)}]^T \quad (12)$$

and $c_{k, \mathcal{T}_k(t)} = \text{col}\{\text{col}\{c_{k\ell, \mathcal{T}_k(t)p}\}_{p \in \mathcal{T}_\ell}\}_{\ell=1}^K$.

From Theorem 1, we can deduce that the steady-state MSD in the estimation of \mathbf{q}_t^o at node k , i.e., $\lim_{i \rightarrow \infty} \|\tilde{\mathbf{q}}_{k,t}^{(i)}\|^2$, can be very large when node k estimates \mathbf{q}_t^o by using a blind convex combination of all the local estimates of all its neighbors. As a result, the attained performance might be worse than the one achieved by the non-cooperative scheme. Note that (11) reduces to zero if $c_{kk, tt} = 1$ for all k and $t \in \mathcal{T}_k$, i.e., the non-cooperative case is indeed unbiased. Next, we will propose a scheme that combines the D-NSPE LMS algorithm and an adaptive multi-task clustering technique to avoid such a performance degradation and still leverage the cooperation among nodes interested in estimating different but overlapping vectors of parameters simultaneously.

4. UNSUPERVISED DIFFUSION-BASED LMS FOR NSPE

Since the nodes do not know a priori the relationship between the NSPE tasks, they can initially focus on solving the NSPE problem of Section 2 by implementing the non-cooperative strategy. In particular, each agent can implement (6)-(8) with $c_{kk, tt}(i) = 1$ for all $k \in \{1, 2, \dots, K\}$ and $t \in \mathcal{T}_k$, which corresponds to the following stand-alone LMS-type adaptation step

$$\mathbf{s}_{k,t}^{(i)} = \mathbf{s}_{k,t}^{(i-1)} + \mu_k \mathbf{U}_{k_t, i}^H \left[\mathbf{d}_{k,i} - \sum_{p \in \mathcal{T}_k} \mathbf{U}_{k_p, i} \mathbf{s}_{k,p}^{(i-1)} \right] \quad (13)$$

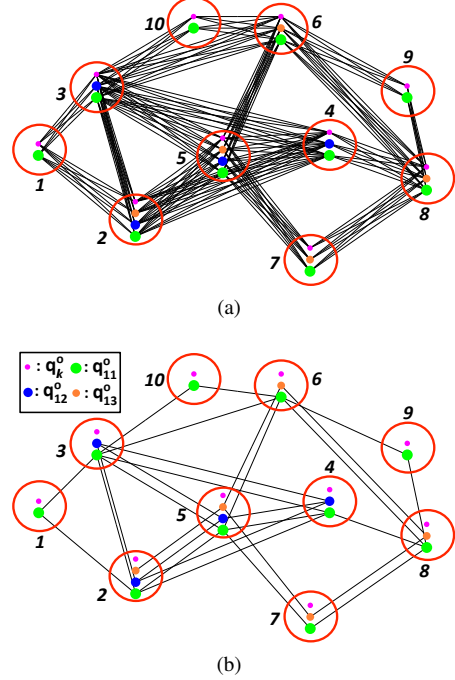


Fig. 2. Topology with all the initial cooperation links (a) and the resulting cooperation links after multi-task clustering (b).

where $\mathbf{s}_{k,t}^{(i)}$ denotes the estimate of \mathbf{q}_t^o resulting from the non-cooperative LMS performed by node k at time instant i with $t \in \mathcal{T}_k$.

Following similar arguments to [22], for sufficiently small step sizes, i.e., $\mu_k \ll 1$, and after sufficient iterations, i.e., $i \rightarrow \infty$, it can be shown that the difference between $\mathbf{s}_{k,t}^{(i)}$ and $\mathbf{s}_{\ell,p}^{(i)}$ follows a Gaussian distribution:

$$\mathbf{s}_{k,t}^{(i)} - \mathbf{s}_{\ell,p}^{(i)} \sim \mathbb{N}(\mathbf{q}_t^o - \mathbf{q}_p^o, \mu_{\max} \Delta_{k\ell, tp}) \quad (14)$$

where $\mu_{\max} = \max_k \mu_k$ and $\Delta_{k\ell, tp}$ is an $M \times M$ symmetric, positive semi-definite matrix with $k, \ell \in \{1, 2, \dots, K\}$, $t \in \mathcal{T}_k$ and $p \in \mathcal{T}_\ell$. If both estimates are associated with the same task, which is denoted as $\mathbf{q}_t^o = \mathbf{q}_p^o$ from now on, from (14) note that with high probability $\|\mathbf{s}_{k,t}^{(i)} - \mathbf{s}_{\ell,p}^{(i)}\|^2 = O(\mu_{\max})$. On the contrary, if the local estimates are associated with different tasks, which will be denoted as $\mathbf{q}_t^o \neq \mathbf{q}_p^o$, then $\|\mathbf{s}_{k,t}^{(i)} - \mathbf{s}_{\ell,p}^{(i)}\|^2 = O(1)$ will hold with high probability. Thus, if node ℓ transmits $\mathbf{s}_{\ell,p}^{(i)}$ to node k with $\ell \in \mathcal{N}_k \setminus \{k\}$ and $p \in \mathcal{T}_\ell$, node k can perform a hypothesis test to determine whether the local estimate $\mathbf{s}_{k,t}^{(i)}$ at node k corresponds to the estimation of the vector of parameters \mathbf{q}_t^o with $t \in \mathcal{T}_k$:

$$\|\mathbf{s}_{k,t}^{(i)} - \mathbf{s}_{\ell,p}^{(i)}\|^2 \underset{\mathbb{H}_1}{\overset{\mathbb{H}_0}{\geq}} \tau_{k\ell, tp} \quad (15)$$

where \mathbb{H}_0 equals the hypothesis $\mathbf{q}_t^o = \mathbf{q}_p^o$, \mathbb{H}_1 denotes the hypothesis $\mathbf{q}_t^o \neq \mathbf{q}_p^o$ and $\tau_{k\ell, tp}$ is a predefined threshold.

A similar reasoning to [22] shows that the probabilities of false alarm and misdetection decay at exponential rates, i.e.,

$$\mathbb{P}[\|\mathbf{s}_{k,t}^{(i)} - \mathbf{s}_{\ell,p}^{(i)}\|^2 > \tau_{k\ell, tp} \mid \mathbf{q}_t^o = \mathbf{q}_p^o] \leq O(e^{-c_1/\mu_{\max}}) \quad (16)$$

$$\mathbb{P}[\|\mathbf{s}_{k,t}^{(i)} - \mathbf{s}_{\ell,p}^{(i)}\|^2 < \tau_{k\ell, tp} \mid \mathbf{q}_t^o \neq \mathbf{q}_p^o] \leq O(e^{-c_2/\mu_{\max}}) \quad (17)$$

where $c_1, c_2 > 0$ denotes some positive constants and $\tau_{k\ell, tp} \in (0, d_{k\ell, tp})$ with $d_{k\ell, tp} = \|q_t^o - q_p^o\|^2$. Thus, if $\tau_{k\ell, tp}$ and μ_{\max} are sufficiently small, as i approaches infinity, node k can know with high probability if a local estimate $\varsigma_{\ell, p}^{(i)}$ at node ℓ is associated with the estimation of q_t^o . This information can be used by node k to simultaneously cluster the local estimates of the nodes according to the task that they aim to solve. In particular, from the following set

$$\mathcal{N}_{k,t}(i) = \left\{ (\ell, p) : \ell \in \mathcal{N}_k, p \in \mathcal{T}_\ell, \|\varsigma_{k,t}^{(i)} - \varsigma_{\ell,p}^{(i)}\|^2 < \tau_{k\ell, tp} \right\} \quad (18)$$

node k can dynamically infer the set of indices $\ell \in \mathcal{N}_\ell$ and $p \in \mathcal{T}_\ell$ for which $c_{k\ell, tp}(i) > 0$ should be verified. At the same time, these task-specific neighborhoods can be used by each node k to perform the following diffusion-based NSPE strategy for each task $t \in \mathcal{T}_k$

$$\begin{cases} \psi_{k,t}^{(i)} = \phi_{k,t}^{(i-1)} + \mu_k \mathbf{U}_{k,t,i}^H \left[\mathbf{d}_{k,i} - \sum_{p \in \mathcal{T}_k} \mathbf{U}_{k,p,i} \phi_{k,p}^{(i-1)} \right] \\ \phi_{k,t}^{(i)} = \sum_{(\ell,p) \in \mathcal{N}_{k,t}(i-1)} c_{k\ell, tp}(i-1) \psi_{\ell,p}^{(i)} \end{cases} \quad (19)$$

where $\sum_{(\ell,p) \in \mathcal{N}_{k,t}(i-1)} c_{k\ell, tp}(i-1) = 1$.

This leads to the following algorithm:

Unsupervised Diffusion-based LMS for NSPE (UD-NSPE)

- Start with any initial guesses $\varsigma_{k,t}^{(0)}$, $\phi_{k,t}^{(0)}$ and $\mathcal{N}_{k,t}(-1) = \{(k, t)\}$ for all $k \in \{1, 2, \dots, K\}$ and $t \in \mathcal{T}_k$.
 - At each time i and each node $k \in \{1, 2, \dots, K\}$:
 1. Update $\varsigma_{k,t}^{(i)}$ by executing (13) for each $t \in \mathcal{T}_k$.
 2. Update $\phi_{k,t}^{(i)}$ by executing recursion (19) over the set $\mathcal{N}_{k,t}(i-1)$ for each $t \in \mathcal{T}_k$.
 3. Update the set $\mathcal{N}_{k,t}(i)$ for each $t \in \mathcal{T}_k$ by using (18) with $\{\varsigma_{\ell,p}^{(i)}; \ell \in \mathcal{N}_k, p \in \mathcal{T}_\ell\}$ from Step 1.
-

Based on the multi-task clustering information resulting from (13) and (18), at each time instant i the proposed UD-NSPE LMS is able to leverage the cooperation among nodes with different but overlapping estimation interests. Despite this fact, from (17), note that UD-NSPE LMS still yields asymptotically unbiased estimates if the $\lim_{i \rightarrow \infty} \mu_k = 0$ for all $k \in \{1, 2, \dots, K\}$.

5. SIMULATIONS

To illustrate the effectiveness of the proposed algorithm, we consider a network formed by $K = 10$ nodes whose initial topology with all possible cooperation links is shown in Fig. 2(a). The red circles denote the nodes and the colored balls inside represent different tasks within each node. In the considered network, each node k is interested in estimating a vector of global parameters $\mathbf{q}_{11}^o \in \mathbb{R}^3$ and a vector of local parameters $\mathbf{q}_k^o \in \mathbb{R}^3$ with $k \in \{1, \dots, K\}$, denoted with green and magenta balls, respectively. Additionally, two different vectors of common parameters coexist, i.e., $\mathbf{q}_{12}^o \in \mathbb{R}^3$ and $\mathbf{q}_{13}^o \in \mathbb{R}^3$, represented by blue and orange balls, respectively. Each entry of the global, common or local vector of parameters is randomly drawn from a uniform distribution defined in the interval $(0, 1)$. Moreover, the data observed by node k follows the observation model given in (1) with $L_k = 1$. Both the background noise $v_{k,i}$ and the regressors $\mathbf{u}_{k,i}$ are independently drawn from a Gaussian

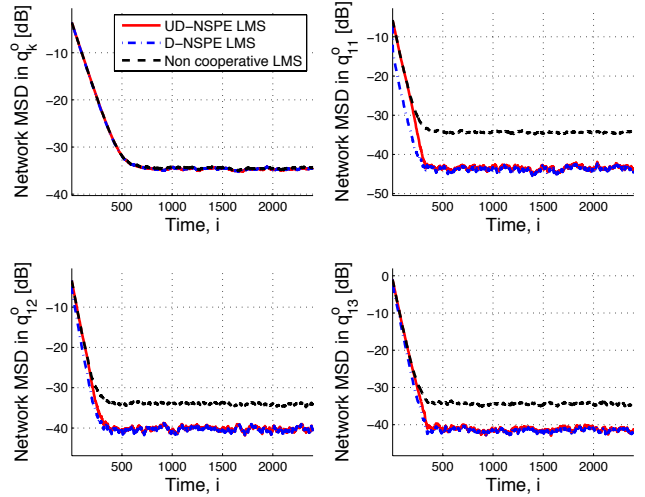


Fig. 3. Temporal evolution of the network MSD for the estimation of the different vectors of parameters.

distribution that is spatially and temporally independent. In particular, $v_{k,i}$ follows a Gaussian distribution with zero mean and variance $\sigma_{v_k}^2 = 10^{-3}$ for all k . Similarly, the regressors $\mathbf{u}_{k,i}$ are zero mean $(1 \times 3n_k)$ random vectors governed by a Gaussian distribution with zero mean and covariance matrix $\mathbf{R}_{\mathbf{u}_{k,i}} = \sigma_{\mathbf{u}_k}^2 \mathbf{I}_{3n_k}$. The variance $\sigma_{\mathbf{u}_k}^2$ is randomly chosen in $(0, 1)$ so that the Signal-to-Noise-Ratio (SNR) at each node ranges from 10 dB to 20 dB. Furthermore, each step-size μ_k is set equal to $4 \cdot 10^{-3}$ and a uniform combination policy has been selected to generate the combination coefficients.

Fig. 2(b) shows the estimated cooperation links in the steady-state for one of the experiments. Note that cooperation links between local estimates of the same vector of parameters at different neighboring nodes remain active. On the contrary, cooperation links between local estimates of different vectors of parameters are dropped. For the UD-NSPE algorithm, the D-NSPE algorithm derived in [16] and the non-cooperative LMS, Fig. 3 plots the learning behaviour in terms of the network MSD associated with the estimation of the vectors of global, common and local parameters. To generate each plot, we have averaged the results over 100 randomly initialized independent experiments. Since the multi-task clustering technique correctly determines the links between local estimates of the same vector of parameters at different neighboring nodes, the UD-NSPE algorithm outperforms the non-cooperative LMS and achieves the same steady-state network MSD as the D-NSPE algorithm.

6. CONCLUSION

We have considered an NSPE problem where the nodes simultaneously estimate vectors of local, common and global interest in a setting where the nodes do not know a priori which of the local estimates of their neighbors correspond to which estimation task. To solve this problem, we have presented an algorithm based on a diffusion-based NSPE LMS and a multi-task clustering technique that lets each node infer which of the local estimates of its neighbors correspond to each of its own estimation tasks. Unlike the existing schemes, the proposed algorithm can yield unbiased estimates for the NSPE problem while still leveraging the cooperation among nodes with different interests. Finally, the effectiveness of the proposed algorithm has been illustrated through computer simulations.

7. REFERENCES

- [1] G. Mateos, I. D. Schizas, and G. B. Giannakis, "Distributed recursive least-squares for consensus-based in-network adaptive estimation," *IEEE Transactions on Signal Processing*, vol. 57, no. 11, pp. 4583–4588, 2009.
- [2] A. G. Dimakis, S. Kar, J. M. F. Moura, M. G. Rabbat, and A. Scaglione, "Gossip algorithms for distributed signal processing," *Proceedings of the IEEE*, vol. 98, no. 11, pp. 1847–1864, 2010.
- [3] C. G. Lopes and A. H. Sayed, "Incremental adaptive strategies over distributed networks," *IEEE Transactions on Signal Processing*, vol. 55, no. 8, pp. 4064–4077, 2007.
- [4] F. S. Cattivelli and A. H. Sayed, "Diffusion LMS strategies for distributed estimation," *IEEE Transactions on Signal Processing*, vol. 58, no. 3, pp. 1035–1048, 2010.
- [5] S. Chouvardas, K. Slavakis, and S. Theodoridis, "Adaptive robust distributed learning in diffusion sensor networks," *IEEE Transactions on Signal Processing*, vol. 59, no. 10, pp. 4692–4707, 2011.
- [6] S. Doclo, M. Moonen, T. Van den Bogaert, and J. Wouters, "Reduced-bandwidth and distributed mwf-based noise reduction algorithms for binaural hearing aids," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 17, no. 1, pp. 38–51, 2009.
- [7] A. Bertrand and M. Moonen, "Distributed adaptive node-specific signal estimation in fully connected sensor networks - part I: Sequential node updating," *IEEE Transactions on Signal Processing*, vol. 58, no. 10, pp. 5277–5291, 2010.
- [8] J. Plata-Chaves, A. Bertrand, and M. Moonen, "Distributed signal estimation in a wireless sensor network with partially-overlapping node-specific interests or source observability," in *IEEE 40th International Conference on Acoustics, Speech and Signal Processing, 2015. ICASSP 2015*, 2015.
- [9] A. Bertrand and M. Moonen, "Distributed node-specific LCMV beamforming in wireless sensor networks," *IEEE Transactions on Signal Processing*, vol. 60, no. 1, pp. 233–246, 2012.
- [10] P. Di Lorenzo, S. Barbarossa, and A. H. Sayed, "Bio-inspired swarming for dynamic radio access based on diffusion adaptation," in *IEEE 19th European Signal Conference, 2011. EUSIPCO 2011*, 2012, pp. 1–6.
- [11] P. Di Lorenzo, S. Barbarossa, and A. H. Sayed, "Bio-inspired decentralized radio access based on swarming mechanisms over adaptive networks," *IEEE Transactions on Signal Processing*, vol. 61, no. 12, pp. 3183–3197, 2013.
- [12] V. Kekatos and G. B. Giannakis, "Distributed robust power system state estimation," *IEEE Transactions on Power Systems*, vol. 28, no. 2, pp. 1617–1626, 2013.
- [13] J. Plata-Chaves, N. Bogdanovic, and K. Berberidis, "Distributed incremental-based RLS for node-specific parameter estimation over adaptive networks," in *IEEE 21st European Signal Conference, 2013. EUSIPCO 2013*, 2013.
- [14] N. Bogdanovic, J. Plata-Chaves, and K. Berberidis, "Distributed incremental-based LMS for node-specific adaptive parameter estimation," *IEEE Transactions on Signal Processing*, vol. 62, no. 20, pp. 5382–5397, 2014.
- [15] J. Chen, C. Richard, A. O. Hero III, and A. H. Sayed, "Diffusion LMS for multitask problems with overlapping hypothesis subspaces," in *IEEE International Workshop on Machine Learning for Signal Processing, MLSP 2014*, 2014.
- [16] J. Plata-Chaves, N. Bogdanovic, and K. Berberidis, "Distributed diffusion-based LMS for node-specific parameter estimation over adaptive networks," *IEEE Transactions on Signal Processing*, vol. 13, no. 63, pp. 3448–3460, 2015.
- [17] J. Chen, C. Richard, and A. H. Sayed, "Multitask diffusion adaptation over networks," *Submitted to Transactions on Signal Processing*, 2013 [Online]. Available: <http://arxiv.org/abs/1311.4894>.
- [18] R. Nassif, C. Richard, A. Ferrari, and Ali H Sayed, "Proximal multitask learning over networks with sparsity-inducing coregularization," *Submitted to IEEE Transactions on Signal Processing*, 2015 [Online]. Available: <http://arxiv.org/abs/1509.01360>.
- [19] J. Chen and A. H. Sayed, "Distributed Pareto-optimal solutions via diffusion adaptation," in *IEEE Statistical Signal Processing Workshop, 2012. SSP 2012.*, 2012, pp. 648–651.
- [20] J. Chen, C. Richard, and A. H. Sayed, "Diffusion LMS over multitask networks," *IEEE Transactions on Signal Processing*, vol. 63, no. 11, pp. 2733–2748, 2015.
- [21] X. Zhao and A. H. Sayed, "Clustering via diffusion adaptation over networks," in *3rd International Workshop on Cognitive Information Processing, 2012. (CIP 2012)*, 2012, pp. 1–6.
- [22] X. Zhao and A. H. Sayed, "Distributed clustering and learning over networks," *IEEE Transactions on Signal Processing*, vol. 63, no. 13, pp. 3285–3300, 2015.
- [23] J. Chen, C. Richard, and Sayed A. H., "Adaptive clustering for multi-task diffusion networks," in *IEEE 23rd European Signal Conference, 2015. EUSIPCO 2015*, 2015, pp. 2746–2750.
- [24] S. Khawatmi, A. M. Zoubir, and Sayed A. H., "Decentralized clustering over adaptive networks," in *IEEE 23rd European Signal Conference, 2015. EUSIPCO 2015*, 2015, pp. 2746–2750.
- [25] A. H. Sayed, "Diffusion adaptation over networks," *To appear in E-Reference Signal Processing, R. Chellapa and S. Theodoridis, Eds., Elsevier, 2013*, 2012 [Online]. Available: <http://arxiv.org/abs/1205.4220>.